Smart Agricultural Technology 12 (2025) 101042

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

Smart and accurate: A new tool to identify stressed soybean seeds based on multispectral images and machine learning models

Ana Carolina Picinini Petronilio [a,*], Clíssia Barboza Mastrangelo [b], Thiago Barbosa Batista [b], Gustavo Roberto Fonseca de Oliveira [a], Isabela Lopes dos Santos [a], Edvaldo Aparecido Amaral da Silva [a]

[a] Department of Crop Sciences, School of Agriculture, São Paulo State University, Botucatu, SP, 18610-034, Brazil
[b] Laboratory of Radiobiology and Environment, Center for Nuclear Energy in Agriculture, University of São Paulo, Piracicaba, SP, 13416-000, Brazil

ARTICLE INFO

Keywords:
Environmental stress
Seed quality
Green seed problem
Chlorophyll fluorescence
Reflectance
Seed physical properties

ABSTRACT

Extreme environmental conditions have been recurrent during the last few years and have impacted crop seed quality worldwide, mainly but not limited to, soybeans (*Glycine* max (L) Merrill). To overcome this, seed companies often demand innovative tools to address seed quality factors. Machine learning models based on multispectral imaging are a novel seed quality analysis approach. Thus, we hypothesize that segmenting stressed (those produced under conditions that are not favorable to the mother-plant) and non-stressed (produced under conditions favorable to the mother-plant) soybean seeds would be possible with this technology, opening a new opportunity for seed quality management and elucidating quality factors. Soybean seeds (cultivar BR/MG 46-Conquista) were produced under water deficit and heat during maturation (from R5.5 onwards). Multispectral images were acquired from stressed and non-stressed seeds, and the reflectance, autofluorescence, physical properties, and chlorophyll parameters were extracted from the images. In parallel, we determined seed vigor. We designed machine learning models using multispectral imaging data based on three algorithms: neural network, support vector machine, and random forest. Our results demonstrated that the stressed seeds have spectral markers that enable their recognition. Concomitantly, these markers had a direct relationship with seed vigor. The machine learning models developed based on neural network algorithm showed the highest performance in segmenting stressed seeds (≥90 % of accuracy, precision, recall, specificity and F1 score) in contrast to random forest- and support vector machine algorithm (≥88 % of accuracy, precision, recall, specificity and F1 score). Here, we report a new approach for multispectral imaging with the potential to identify soybean seeds of lower vigor as a result of unfavorable environmental conditions during seed maturation.

1. Introduction

Soybean is a significant source of protein and oil and is utilized in several industrial products [12]. Due to the many applications of this crop, a strong market has developed worldwide [36]. Many inputs are required to boost the soybean industry, and one of these inputs is seeds. Seeds are responsible for establishing a new set of plants in the field. High-quality seeds are primordial to providing uniform plant development that favors maximum crop productivity [24]. To ensure high-quality seeds, seed companies strive and constantly seek innovations to overcome seed quality challenges.

One of the potential new approaches to be used in seed quality management is multispectral imaging associated with artificial intelligence. Multispectral imaging consists of reflectance or autofluorescence images obtained through the emission of several light wavelengths [19]. This technique can provide optical markers regarding seed chemical composition, pigments, and physical traits such as texture, area, length, width, color, and coat brightness [9,18]. Intact seeds in the samples provide precise data since the image pixels are transformed into numbers through an algorithm, and consequently eliminate the subjectivity of human interpretation, typical of traditional seed quality tests (e.g., germination, accelerated aging test and tetrazolium) [8,19]. The data generated combined with machine learning (ML) models allow the use of this technique in seed processing to evaluate their quality autonomously [16].

Multispectral imaging has recently been the object of study to

* Corresponding author.
E-mail address: anacpetronilio@gmail.com (A.C.P. Petronilio).

https://doi.org/10.1016/j.atech.2025.101042
Received 27 March 2025; Received in revised form 19 May 2025; Accepted 21 May 2025
Available online 22 May 2025

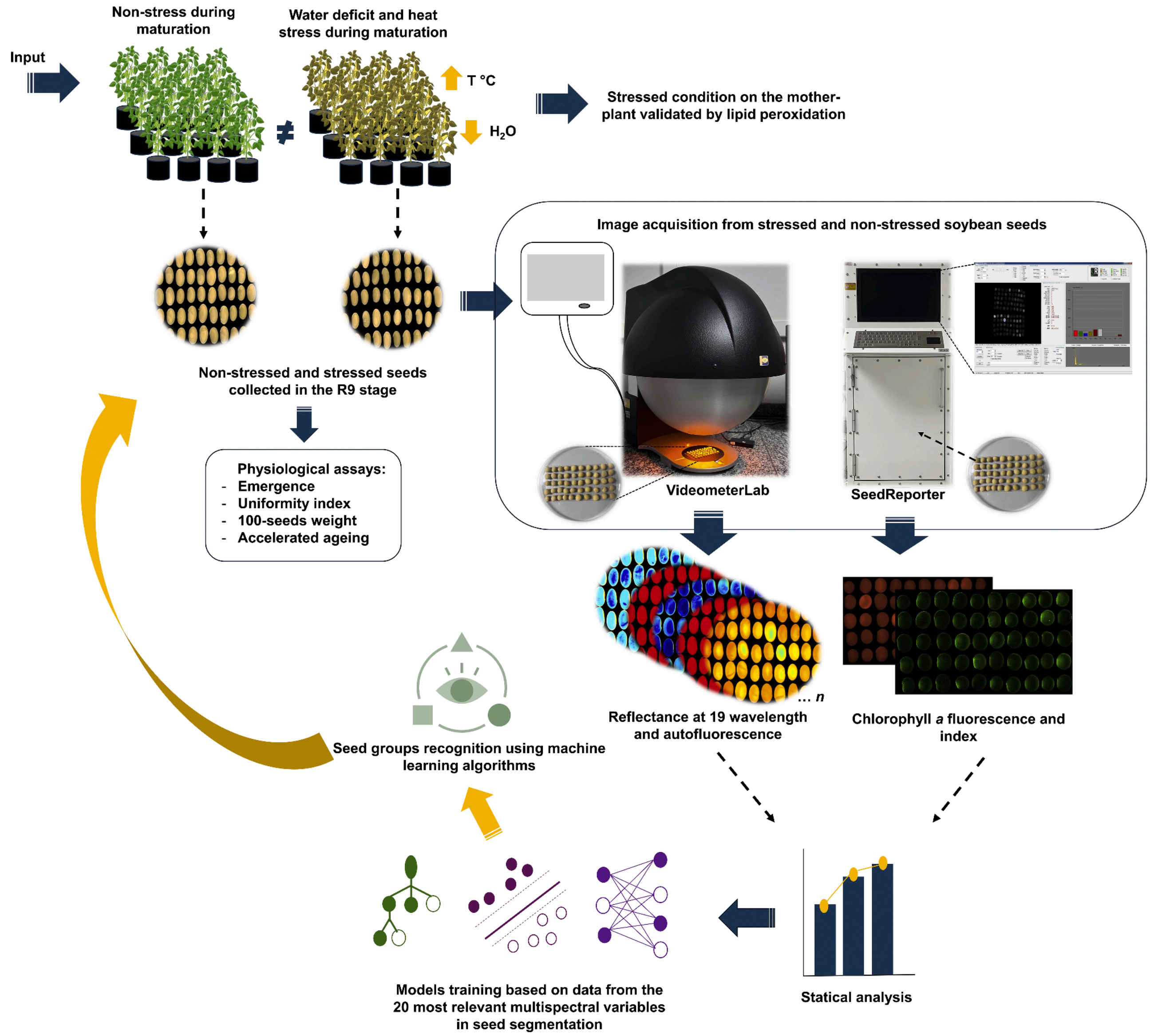

Fig. 1. Framework of the main procedure to produce soybean seeds and capture and analyze multispectral imaging.

develop seed quality management programs. For example, Barboza da Silva et al. [7] showed that autofluorescence signals from the 365/400 nm wavelength associated with ML models could identify artificially aged soybean seeds. Chlorophyll (Chl) fluorescence, anthocyanin, 660, 690, and 780 nm light reflectance, and physical attributes associated with ML models were used to select high-quality peanut seeds [16]. Autofluorescence-spectral imaging related to Chl was used to segment soybean seeds in different maturation stages [8]. This technique was also used to identify viability, vigor, and hard seeds from *Medicago sativa* [38,39]. Texture, color, and 490 nm reflectance associated with artificial intelligence tools were used to discriminate pathogenic fungi in peanut seeds [34].

The soybean seed market deals with several aspects affecting seed quality. In recent decades, reports show that soybean seeds have germination, vigor, and longevity impacted due to abiotic environmental stress [1,8,35]. One of the reported stresses that impact seeds is a combination of water deficit and heat [30,35], and the occurrence of this stress during the plant maturation phase was outlined as being more critical than during the vegetative or embryogenic phases [11]. Environmental stress negatively impacts seed physiological quality [1,41] and is one of the challenges seed companies must overcome to ensure a high-quality product. Therefore, a tool that enables seed companies to know the causes of seed quality issues quickly and non-destructively would be helpful for seed quality management.

We hypothesize that there are morpho(physiological) changes on seeds induced by environmental stress over the mother-plant which can be mapped and consequently provide an identity to stressed soybean seeds. Therefore, it would be possible to promptly identify reductions in seed quality associated with environmental stress. To test this, we used an approach that combined multispectral imaging features and ML models (neural network, random forest and support vector machine) could effectively recognize stressed soybean seeds.

2. Material and methods

We design a framework to summarize the man points to develop this study related the methodology according to Fig. 1.

2.1. Plant material

Soybean seeds (cultivar BR/MG 46-Conquista) were propagated in a greenhouse at the School of Agriculture, Department of Crop Sciences,

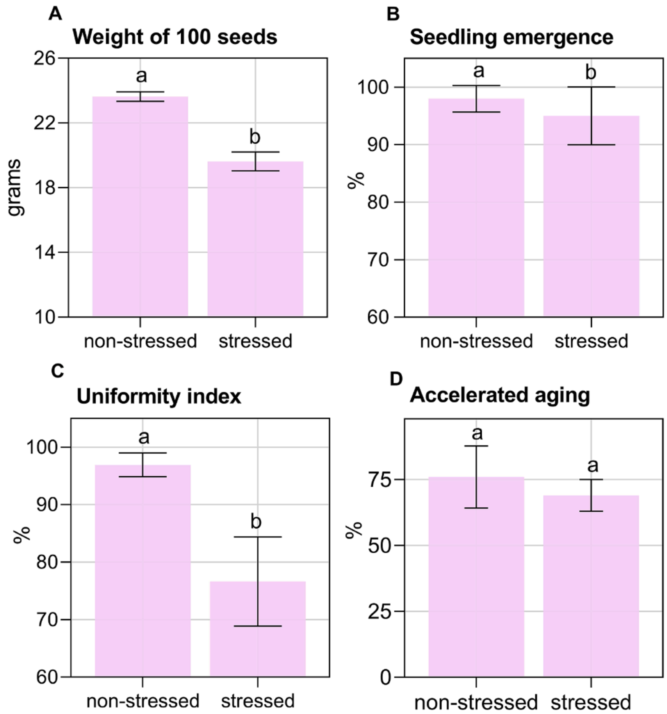

Fig. 2. Effect of environmental stress during maturation of soybean on seed weight and vigor. Seed weight (A), seedling emergence (B), emergence uniformity index (C), and accelerated aging (D) of non-stressed and stressed soybean seeds. Different letters indicate a significant difference ($P \leq 0.05$) by *t*-test. Error bars show the standard deviation from four samples.

UNESP/Botucatu/São Paulo/Brazil (Altitude 810 m). Supplementary lighting was provided through LED full spectrum lamps (28 W), with photosynthetically active radiation emission of 150 µmols m^{-2} s^{-1}. The plants were cultivated in 11 L pots in soil with a texture of 778 g kg^{-1} of sand. The adjustment in the soil acidity and the amount of nutrients applied (nitrogen, phosphorus, potassium and boron) were based on the previous chemical analysis of the soil using the recommendations for soybeans described in Ambrosano et al. [3].The mean temperature during crop development was 28 ± 1 °C during the day and 21 $^{\circ}C \pm 2$ at night. The humidity of the pots was kept close to the field capacity (12 kPa), which had been previously determined. The humidity was controlled twice a day (10:00 am and 4:00 pm). It was measured with a humidity sensor positioned at a depth of 10 cm (in the medium area of the pot). If there was a difference between the reading and the field capacity humidity, the corresponding water volume was replaced to return the soil to the desired water potential (12 kPa).

The sample groups were composed of different environmental conditions during the seed maturation phase: non-stressful (control) and stressful conditions. From the R5.5 stage [31] onwards, 50 % of the pots were randomly assigned to stressful environmental conditions (stressed group), and the other 50 % were kept in the conditions above and constituted the control group (non-stressed group). The environmental stress condition constituted a daytime average of 33 ± 1 °C and nighttime 25 ± 0.4 °C. Water supplementation was kept to a minimum so as to maintain a water potential of 27 kPa (water deficit) until the R9 stage (full maturation).

On the seventh day, after the plants were subjected to stressful conditions, the third wholly expanded leaf from the top of the plants was collected to characterize their oxidative stress (i. e., hydrogen peroxide (H_2O_2) and lipid peroxidation (malondialdehyde – MDA). The oxidative stress was determined with four replicates per treatment (stressed and non-stressed), constituting one leaf per replicate. The samples were immediately frozen in liquid nitrogen and stored in an ultra-freezer (−80 °C) until the evaluation.

For hydrogen peroxide quantification (H_2O_2), 250 mg of leaf tissue was ground and homogenized in 3 mL of 0.1 % (m/v) trichloroacetic acid (TCA) solution with 20 % PVPP (m/m). The solution was centrifuged at 14,000 g at 4 °C for 20 min (Hettich, Universal 320R, Tuttlingen, Germany). A 0.2 mL aliquot of the supernatant was added to the reaction of 0.2 mL of 100 mM potassium phosphate buffer solution (pH 7.5) and 0.8 mL of potassium iodide (KI) solution 1 M. The solution was incubated for one hour on ice in the dark. Then, the absorbance reading was performed on a spectrophotometer (Shimadzu, UV-2700, Kyoto, Japan) at 390 nm. The H_2O_2 concentration was calculated using the H_2O_2 standard curve at 1000 µmol mL^{-1}, and the results were expressed in mmol g^{-1} MF [2].

For lipid peroxidation (malondialdehyde – MDA), 250 mg of leaf tissue were ground in liquid nitrogen, homogenized in 3 mL of 0.1 % (m/v) trichloroacetic acid (TCA) solution with 20 % PVPP (m/m), and centrifuged at 14,000 g at 4 °C, for 20 min (Hettich, Universal 320R, Tuttlingen, Germany). A 0.5 ml aliquot of the supernatant was added to 1.5 ml of 0.5 % (w/v) 2-thiobarbituric acid (TBA) solution prepared in 20 % TCA (w/v) and incubated at 90 °C for 20 min. The reaction was stopped by putting the tube in an ice bath for 10 min. Afterward, the samples were centrifuged at 10,000 g at 4 °C for 5 min. Absorbances were determined using a spectrophotometer (Shimadzu, UV-2700, Kyoto, Japan) at 532 and 660 nm. The malondialdehyde acid (MDA) concentration was calculated using the molar extinction coefficient of 155 mM^{-1} cm^{-1} and expressed as MDA nmol g^{-1} MF [10].

The seeds were harvested at the R9 stage [31], and unformed seeds were discarded. The seeds were stored in a cold chamber at 12 °C/55 % RH for one week, and then the tests were carried out.

2.2. Seed physiological quality assays

Before carrying out the seed quality tests, the seed water content of the seeds was balanced to 0.15 g H2O/g DW^{-1} by pre-conditioning at 25 °C for 16 hours. The water content was determined according to Ista [23].

Four replicates of 100 seeds were weighed on an analytical scale with a precision of 0.001 g to determine the weight of the seeds. To evaluate the seedling emergence and uniformity index, four replicates of 25 seeds were sown at a depth of 5 cm in 200-cell trays with organo-mineral substrate moistened with 60 % of its water-holding capacity. The seedlings were grown at 25 $^{\circ}C \pm 0.8$. Emerged seedlings were counted daily at the same time until the number of emerged seedlings stabilized. For seedling emergence, data were expressed as a percentage of emerged seedlings an the uniformity index was calculated according to Ebone et al. [14]. For the accelerated aging test (AA), four replicates of 25 seeds were scattered on a stainless-steel screen inside a plastic box (11.0 × 11.0 × 3.5 cm) containing 40 mL of deionized water, maintained at 41 °C for 72 h. Afterward, a germination test was performed according to Ista [23].

2.3. Multispectral imaging data acquisition

250 seeds per group were placed on acetate sheets in the same position (hilum facing left), and the images were captured using a VideometerLab4™ equipment (Videometer A/S, Herlev, Denmark), according to Batista et al. [8]. Reflectance images were acquired at 19 wavelengths: 365, 405, 430, 450, 470, 490, 515, 540, 570, 590, 630, 645, 660, 690, 780, 850, 880, 940 and 970 nm. Autofluorescence images were acquired using 12 different excitation/emission combinations at wavelengths: 365/400, 365/500, 405/500, 430/500, 450/500, 470/500, 365/600, 405/600, 430/600, 450/600, 470/600, 490/600, 515/600, 540/600, 570/600, 365/700, 405/700, 430/700, 450/700, 470/700, 490/700, 515/700, 540/700, 570/700, 590/700, 630/700, 645/ 700, and 660/700 nm [22]. After image acquisition, data were extracted using VideometerLab™ software (version 3.14.9) according to Barboza da Silva et al. [6]. In addition to the reflectance and

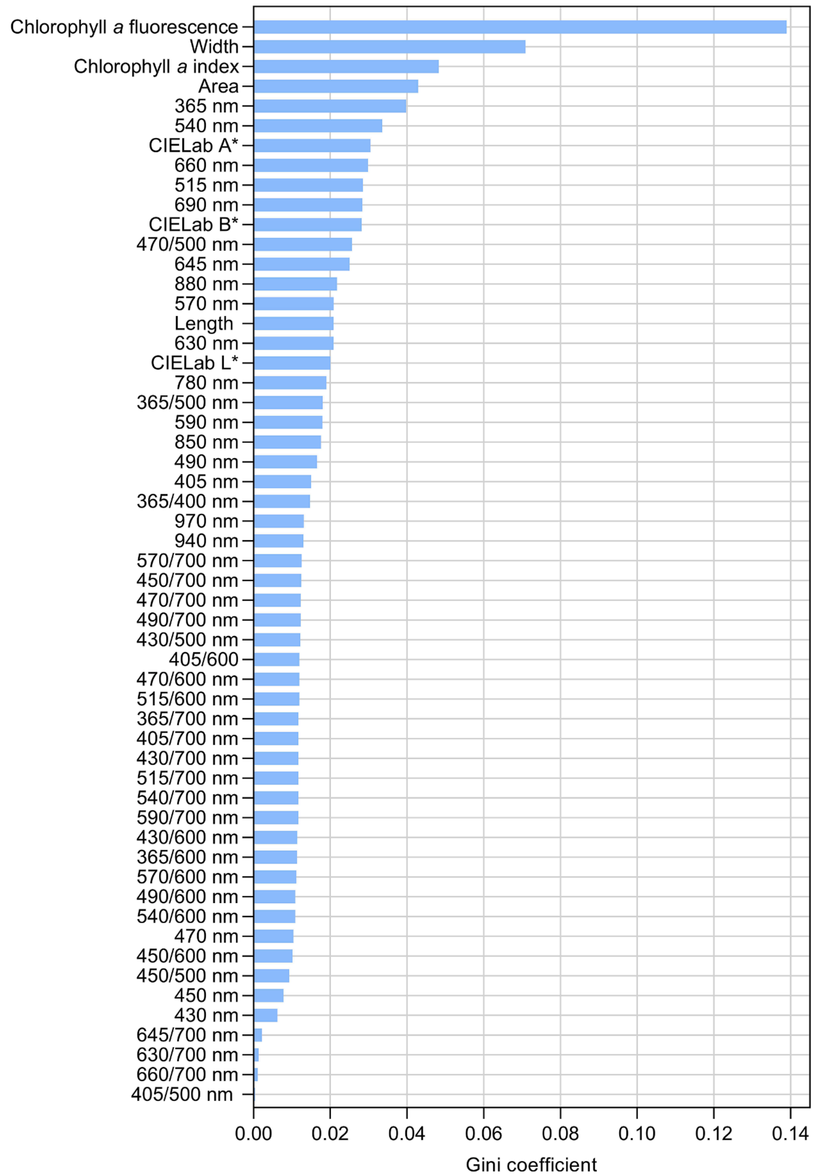

Fig. 3. Gini coefficient for the contribution of the multispectral variables in the segmentation of stressed and non-stressed soybean seeds ($n = 500$ seeds). Chlorophyll *a* fluorescence and the excitation/emission combination of 405/500 nm were the variables that respectively contributed the most and the least to seed group segmentation.

autofluorescence variables, data were determined for the seed physical characteristics: area, length, and width, and for color descriptors: CIELab L^*, CIELab a^*, CIELab b^* [28].

With the same sample preparation, we obtained multispectral images (2448 × 2448 pixels) using a SeedReporter™ (PhenoVation B.V., Wageningen, The Netherlands) instrument. The Chl *a* index images were acquired in the 710 and 770 nm reflectance [21] and the Chl *a* fluorescence was obtained by excitation at 620 nm wavelength and detection at 730 nm [19]. The image acquisition and data extraction were obtained using the SeedReporter™ software version 5.5.1.

2.4. Seed segmentation using machine learning models

We designed ML models using the 20 most important variables of the multispectral data from each single seed ($n = 500$ seeds). The models were based on neural network (NN) using the Multi-layer Perceptron algorithm (solver: Stochastic Gradient Descent; hidden layer sizes: two layers with 25 neurons in each; Activation function: Tanh; Learning rate: adaptive; the maximum number of interactions: 5000) [7], support vector machine algorithm (SVM) (cost: 1; regression loss epsilon [ε] = 0.10; kernel function: radial basis; interaction limit: 100) and random forest algorithm (RF) (number of trees: 10; trees considered at each split: 5; growth control: do not split subsets smaller than 5) [8], in which 70 % of data ($n = 350$) were automatically separated to train the algorithms (K-fold = 5), and 30 % ($n = 150$) were used for the external validation test.

2.5. Statistical design

The physiological ($n = 8$) and multispectral ($n = 500$) data were compared by *t*-test ($P < 0.05$). Additionally, in the multispectral data, we performed (i) Gini coefficient to rank variables regarding their importance in seed group segmentation, (ii) principal component analysis (PCA) to visualize the spatial seed group segregation (stressed and non-stressed) concerning the variables using "ggbiplot" R package. Accuracy, F1, precision, recall, and specificity were used to test the performance of the ML models generated. All metrics were calculated using a confusion matrix based on true Positive (TP), false positive (FP), true

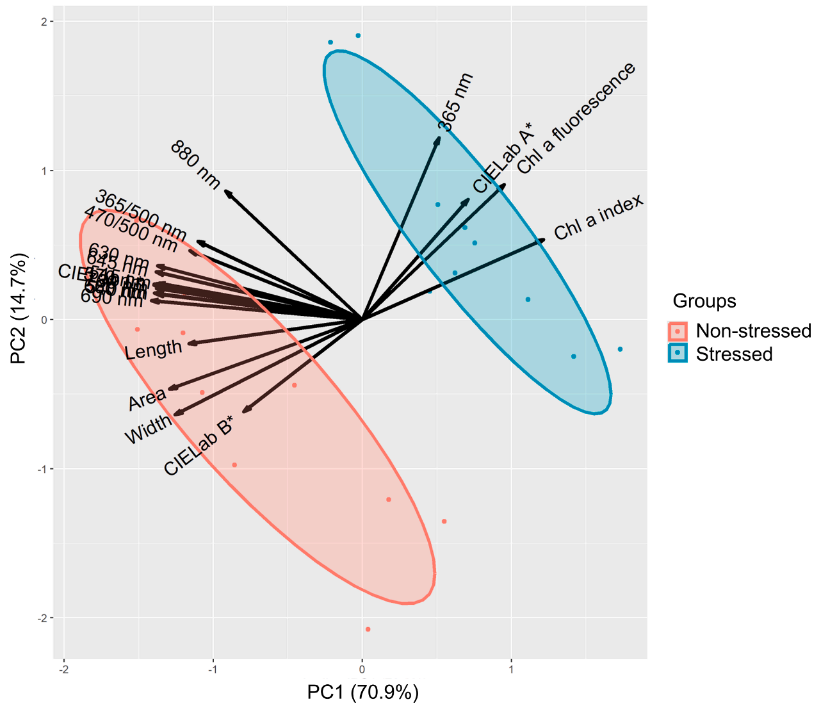

Fig. 4. Biplots of principal component analysis (PCA) for chlorophyll *a* index, chlorophyll *a* fluorescence, physical properties, reflectance, and autofluorescence spectral markers of stressed and non-stressed soybean seeds. The vectors indicate the correlation between the seed groups (stressed and non-stressed) and the dimensions PC1 and PC2.

negative (TN), and false negative (FN) data, where:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1\ score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (4)$$

$$Specificity = \frac{TN}{TN + FP} \quad (5)$$

The Python version 3.11 environment was used to generate the models. Pearson's correlation coefficient was measured to investigate the relationship between the multispectral markers and the seed physiological quality traits using the "corrplot" R package. The adjectives to describe the magnitude of the correlations were proposed by Davis [42], where the values from $r = 0.01$ to 0.09 are negligible correlations, $r =$ 0.10to 0.29 are low, $r = 0.30$ to 0.49 are moderate, $r = 0.50$ to 0.69 are substantial, $r = 0.70$ to 0.99are very high, and $r = 1.0$ is the perfect correlation.

3. Results

3.1. Seed quality characterization

The hydrogen peroxide (H_2O_2) and lipid peroxidation (MDA) content (Fig. S1A, B) increased due to the stressful conditions in soybean plants. Moreover, seed weight was significantly affected by environmental stress. Stressed seeds had less mass than the non-stressed seeds, which was 19.62 g and 23.63 g, respectively (Fig. 2A). Therefore, these results confirmed the effect of plant stress in our research. In parallel, we performed seed physiological quality tests to show seed performance. Stressed seeds had their quality impaired in terms of seedling emergence and emergence uniformity. Stressed seeds had low seedling emergence performance in relation to non-stressed seeds (Fig. 2B and C), with a substantial difference of 20 % in the emergence uniformity, a robust index to measure disturbances in the seed-seedling transition directly related to seed vigor. However, there was no significant difference in the AA test between stressed and non-stressed seeds (Fig. 2D).

3.2. Multispectral parameters can segment stressed seeds

We applied the Gini coefficient to the 55 multispectral variables (reflectance, autofluorescence, physical properties, Chl *a* fluorescence, and Chl *a* index) obtained from the stressed and non-stressed seeds to rank the 20 most important variables. We identified that they are, in order: Chl *a* fluorescence, seed width, Chl *a* index, seed area, 365 nm, 540 nm, CIELab a^*, 660 nm, 515 nm, 690 nm, CIELab b^*, 470/500 nm, 645 nm, 880 nm, 570 nm, seed length, 630 nm, CIELab L^*, 780 nm, 365/500 nm. The entire Gini-coefficient report is shown in Fig. 3.

Using the 20 most important variables based on the Gini coefficient, we applied a PCA to explore seed segmentation (Fig. 4). Component 1 (PC1) and component 2 (PC2) explained 70.9 % and 14.7 % of the significant variation, respectively. PCA revealed a clustering of variables according to the seed production environment (stressed and non-stressed). Stressed seeds were associated with Chl *a* index, Chl *a* fluorescence, CIELab a^*, and 365 nm reflectance. Meanwhile, the other 16 most important variables were associated with non-stressed seeds.

Fig. 5 shows the means of the five most important markers for segmenting stressed and non-stressed seeds based on the Gini coefficient. Stressed seeds had higher values of Chl *a* fluorescence (Fig. 5A), Chl *a* index (Fig. 5B), and reflectance in the 365 nm wavelength (Fig. 5E) ($P \leq 0.05$), as also shown in the images (Fig. 6). However, they had smaller seed width (Fig. 5C) and seed area (Fig. 5D) compared to non-stressed seeds.

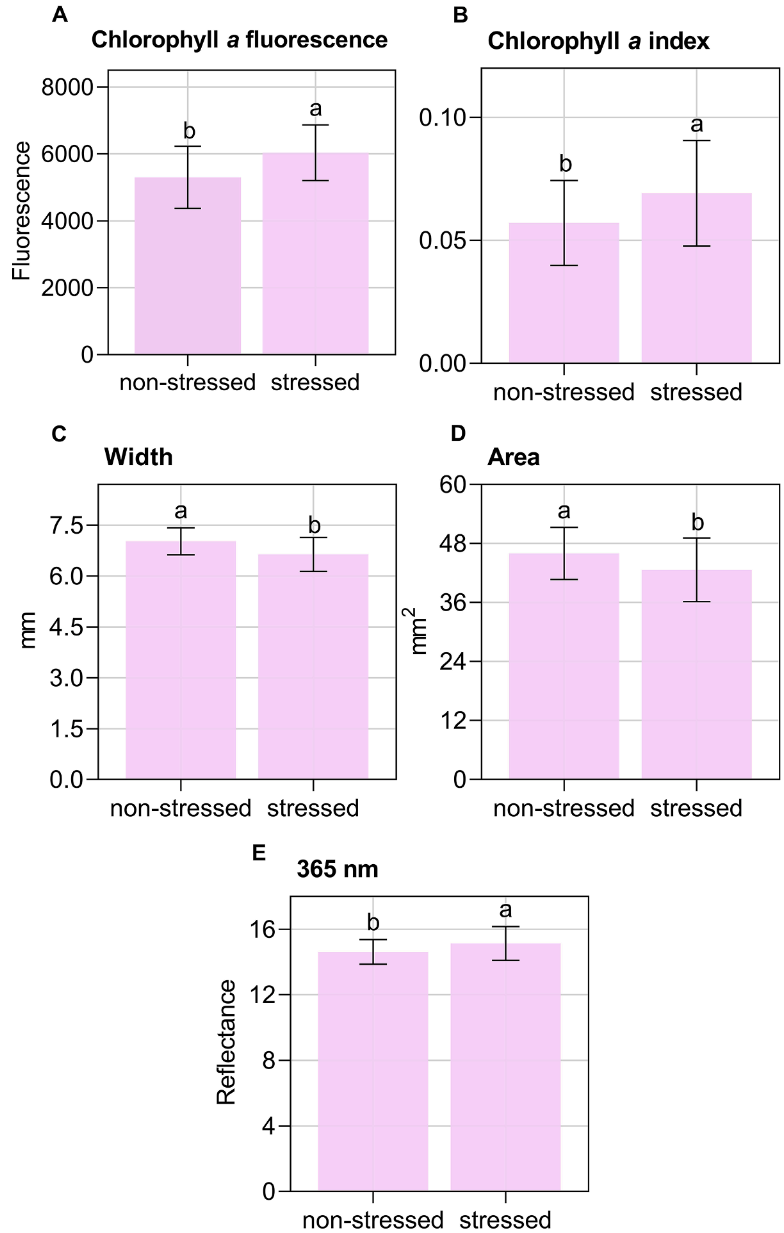

Fig. 5. The five most important multispectral variables to separate non-stressed seeds from stressed soybean seeds based on the Gini coefficient. Different letters indicate a significant difference ($P \leq 0.05$) by *t*-test. Error bars show the standard deviation from 250 samples.

3.3. *Seed segmentation based on different machine learning models*

Using NN as an algorithm, we obtained accuracy, F1, precision, recall, and specificity ≥ 0.92, while RF and SVM presented ≤ 0.88 in the training of the algorithms. In the test, the NN algorithm showed metrics ≥ 0.90; on the other hand, RF and SVM showed metrics ≤ 0.82 (Table 1). In addition, the confusion matrix showed that using NN algorithm in the multispectral data, the probability of success in the prediction is greater than 90 %. The probability of success for stressed seeds identification was 94 %, demonstrating a high capacity for prediction through the parameters used to identify this class (Fig. 7).

3.4. *Correlation of the seed physiological traits with Chl a parameters, physical markers, reflectance, and autofluorescence*

The 20 most important multispectral markers were inserted in a correlation matrix with the physiological data to check the relationship between the variables (Fig. 8). The weight of 100 seeds was strongly correlated with Chl *a* index (−0.95). There was also a very high correlation coefficient between the weight of 100 seeds and physical properties, especially seed width (0.92). Furthermore, the weight of 100 seeds and reflectance obtained a very high correlation coefficient with 365 nm (−0.92) wavelength.

There was a very high correlation coefficient between the seedling emergence and seed area (0.70). The uniformity index was negatively correlated with Chl *a* parameters: Chl *a* index (−0.88) and Chl *a* fluorescence (−0.77). The uniformity index was also very high correlated with physical markers, in which the highest correlation coefficient was obtained with seed width (0.89). Finally, the uniformity index was mainly correlated with the wavelength of 365 nm (−0.83).

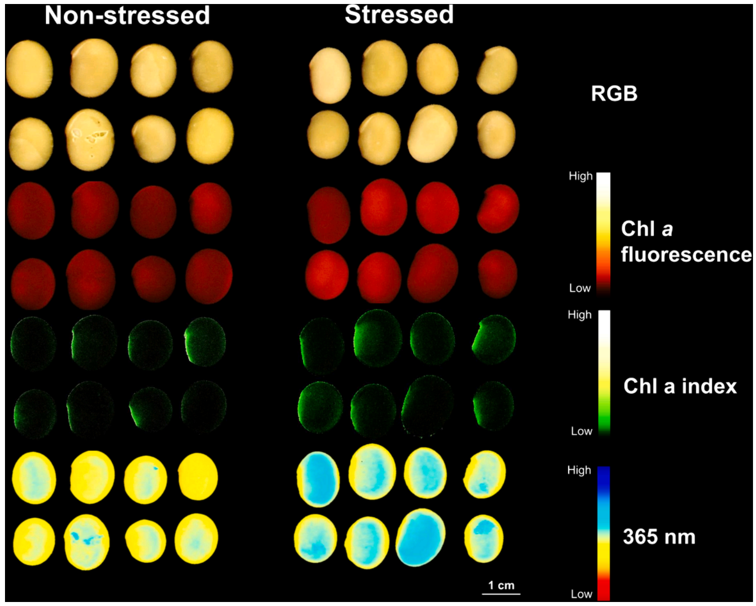

Fig. 6. Score pixels of non-stressed and stressed soybean seed images. RGB images (red, green and blue channels) and corresponding chlorophyll *a* fluorescence (Chl *a* fluorescence), chlorophyll *a* index (Chl *a* index), and reflectance images at 365 nm.

Table 1
Machine learning models to discriminate stressed and non-stressed soybean seeds using multispectral data based on Random Forest (RF), Support Vector Machine (SVM) and Neural Network (NN) algorithms. Models' performance based on accuracy, F1, precision, recall, and specificity.

Metrics	Neural Network		Randon Forest		Suport Vector Machine	
	Training (n = 350)	Test (n = 150)	Training (n = 350)	Test (n = 150)	Training (n = 350)	Test (n = 150)
Acuraccy	0.92	0.91	0.87	0.82	0.81	0.78
F1	0.92	0.90	0.88	0.81	0.82	0.77
Precision	0.92	0.90	0.88	0.82	0.82	0.78
Recall	0.92	0.90	0.87	0.82	0.81	0.78
specificity	0.92	0.90	0.87	0.82	0.81	0.78

4. Discussion

Seed companies require seed quality management tools. Multispectral imaging associated with ML algorithms is a prospective innovation in this field. We asked ourselves if this imaging technology could identify seeds that went through a stressful production environment and then if ML algorithms could automatically segregate the stressed seeds. For this, stressed soybean seeds (seeds produced under stressful conditions of water deficit and heat) and non-stressed soybean seeds – our control group - were analyzed for physiological quality, and also using multispectral imaging tools. Here, we demonstrate for the first time that by using emerging technologies, it is possible to classify the soybean seeds' origin environment and, thus, infer the consequences on seed physiological quality.

In croplands worldwide, sowing seeds with high physiological quality is essential, because it provides fast and uniform germination as demonstrated for soybean seeds by Ebone et al. [14]. Our data showed a reduction in the seed vigor for stressed seeds, as is also reported in the literature [5,15], which causes a disturbance in the seed-seedling transition, reducing emergence uniformity (Fig. 2C). In the fields, uneven seedling emergence produces dominated plants in the plant community, which although being able to recover their height in the late growth stages due to phenotypic plasticity, are less productive because of the reduction in the number of nodes and pods per plant [14]. Reduced seedling emergence rate (Fig. 2B) causes missing plants in the plant stand, directly reducing yield [24].

To test and determine seed vigor, seed companies often apply AA tests in their lab work routine [17]. In our study, the AA test (72 h 41 °C) could not detect seed vigor difference between the two groups (stressed and non-stressed) (Fig. 2D). Thus, seeds with reduced vigor due to stressful conditions during the maturation phase might not be detected by traditional tests, such as AA. This problem is typical because: (i) the reduction in vigor is not as apparent as other factors that affect quality with a latent or immediate effect (such as mechanical damage) and (ii) immediately after harvest, seed vigor can be very high, and differences are not detected in primary testing.

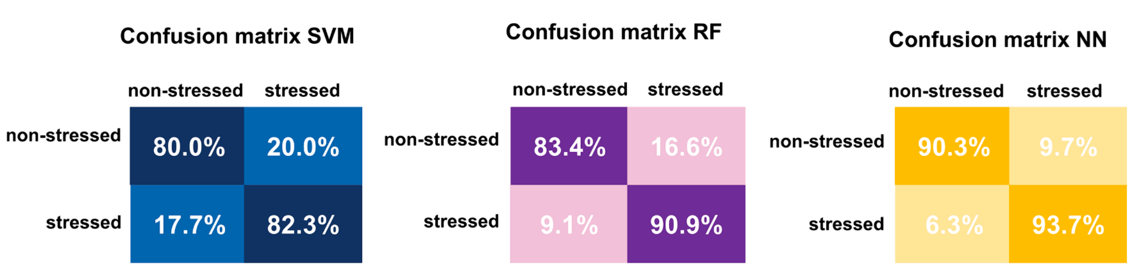

Fig. 7. Confusion matrices to discriminate stressed and non-stressed soybean seeds using multispectral data based on Random Forest (RF), Support Vector Machine (SVM) and Neural Network (NN) algorithms. Confusion matrices were created using the test dataset (n = 150 seeds).

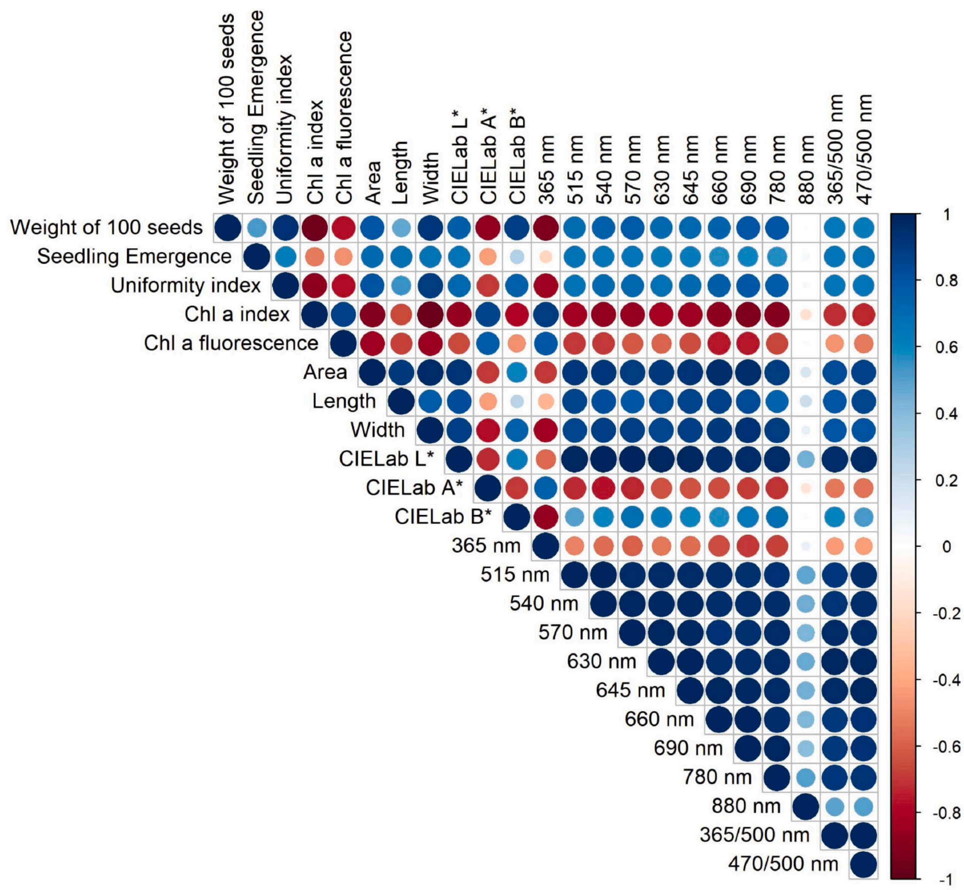

Fig. 8. Pearson's correlation coefficients between seed physiological quality, physical properties, reflectance, and autofluorescence-spectral markers (nm) in stressed and non-stressed soybean seeds.

We achieved an innovative methodology to diagnose reduction in seed vigor caused by stressful environments using Chl *a* fluorescence, Chl *a* index, seed width, seed area, reflectance variables, etc. since the changes in these properties promoted by stress during seed maturation were detected using multispectral imaging (Figs. 3, 5, and 6). Stressed seeds presented high Chl levels, as shown by Chl *a* fluorescence and Chl *a* index (Fig. 5A, 5B and 6). Stressful conditions during maturation promote Chl retention in soybean seeds [29,35] that impact the seed-seedling transition [1]. Thus, Chl levels in stressed seeds were higher, which became a strong marker detected by multispectral images. This technique has high sensitivity and enables distinguishing groups based on the level of chlorophyll that does not degrade due to stressful conditions.

In our study, the greenish color (high chlorophyll content) of the seeds was not visible to the human eye as shown in the RGB images (Fig. 6). High chlorophyll levels impact legume seeds' lifespan [27,40], but the reduction in the physiological quality of soybean seeds is generally detected only at the time of the effective use of the seeds, i.e., after six months of storage waiting to be commercialized. Thus, our method based on multispectral imaging, incorporating data related to chlorophyll fluorescence, can be used as a tool for predicting severe problems that seeds may face after sewing (i.e., loss of viability and consequent commercial standard loss).

In seeds, several fluorescent chemical components can be detected by multispectral images at short wavelengths [13,20]. Fluorescent compounds, such as chlorophyll, can alter the color properties of seeds, especially those related to brightness ([16,28,34]. In parallel, these changes potentially influence the reflectance properties of seeds, as shown in our study at 365 nm (Figs. 5 and 6). This also occurred in peanut [16], tomato and carrot [19]. Therefore, the environmental stress conditions imposed on soybean plants generated seeds with a detectable post-harvest spectral pattern. This spectral pattern opens up a technological opportunity for assisted seed lot selection based on reflectance, physical properties, and color descriptors.

Regarding the shape markers, width and area were the most important variables for segmenting stressed seeds (Fig. 3). Probably, stressed seeds had smaller width, area (Fig. 5C and D), and weight (Fig. 2A) because the plant needed to use its photoassimilates to deal with the oxidative stress (Fig. S1), and the accumulation of reserves in the seed was reduced [32,33]. Thus, the shape of the seeds was modified, which is a common effect in seeds produced over stressful conditions. Integrating shape properties with the chlorophyll fluorescence and reflectance reported here allows us to generate a robust model to classify seeds that go through stress during their development (Table 1 and Fig. 7).

Additionally, we demonstrated that our data could be interpreted using machine learning algorithms, especially NN (Table 1 and Fig. 7). This opens the possibility to turn our model into a "results manager," i.e., identifying why there is a reduction in physiological quality in seeds when there is no noticeable damage, the so-called "silent enemies", as residual chlorophylls that impact the seed-seedling transition as reported by Ajala-luccas et al. [1] and observed in our data (Figs. 2 and 5).

In another way, our results allow more precise and sensitive diagnoses since the technology is highly associated with the uniformity index (Fig. 8), an effective vigor test diagnosis. In this context, Barboza da Silva et al. [7] developed markers based on autofluorescence-spectral combined with ML to classify soybean seeds regarding their vigor. Therefore, our results additionally demonstrate the potential of these tools in this matter.

The technology proposed here can also be used in plant breeding studies. With global climate change, the tool can help to identify stress-tolerant materials by screening different genotypes submitted to stress. It can also be used in stress memory tests. Like phenotypic plasticity, stress memory is a plant survival strategy [4]. The so-called plant stress memory is produced depending on the stress type and level since plants can carry out epigenetic modifications that will result in changes in their metabolism to ensure adaptation to the new environment [26]. This change can be passed on to the following generations and produce adaptations to new environmental scenarios [25]. Stress memory assessments have been used in plant breeding programs to select stress-resistant materials [37]. The great advantage of adapting the tool in memory stress assays to seed bulking is that it allows quantification on a single seed.

Our study brings more knowledge about soybean seed quality evaluation through multispectral imaging associated with ML algorithms and presents a new approach to detect vigor differences caused by stress occurrence during seed maturation in a much more precise way than the traditional seed physiological quality tests.

5. Conclusion

Soybean seeds produced under stressful conditions for the mother-plant have their physiological quality reduced, especially noted by the 20 % reduction in the uniformity index, directly related to the establishment of seedlings.

Using multispectral images, it is possible to map several parameters that change as a function of seed stress, among which the five most important parameters are: chlorophyll *a* fluorescence, chlorophyll *a* index, and reflectance at 365 nm (which increase in stressed seeds); area and width (which reduce in stressed seeds).

Our approach using multispectral imaging combined with machine learning was efficient in classify and identify seeds according to the environmental condition of the production (i.e., non-stressed and stressed seeds), in particular, using neural networks, which achieved learning and recognition of seeds in the different classes above 90 %.

Funding sources

This work was supported by São Paulo Research Foundation–FAPESP [grant numbers 2017/15220–7; 2018/03793–5; 2018/03802–4; 2018/01774–3, 2020/14050–3, 2023/00435–9, 2019/06160–6, 2021/03209–4, 2017/50211–9s], National Council for Scientific and Technological Development – CNPq [grant numbers 314305/2021–1, 142236/2020–9, 311526/2021–7] and Coordination for the Improvement of Higher Education Personnel – Brazil – CAPES [grant numbers 88887.670559/2022–00; 88887.685215/2022–00; 88887.924553/2023–00].

CRediT authorship contribution statement

Ana Carolina Picinini Petronilio: Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Clíssia Barboza Mastrangelo:** Writing – review & editing, Resources, Investigation, Funding acquisition, Conceptualization. **Thiago Barbosa Batista:** Writing – review & editing, Visualization, Investigation, Formal analysis, Data curation. **Gustavo Roberto Fonseca de Oliveira:** Writing – review & editing, Investigation, Data curation. **Isabela Lopes dos Santos:** Methodology, Investigation. **Edvaldo Aparecido Amaral da Silva:** Writing – review & editing, Supervision, Resources, Funding acquisition, Conceptualization.

Declaration of competing interest

None.

Acknowledgments

We are thankful to Valeria Cristina Retameiro Giandoni and Dennis Vinícius Lopes de Souza for their support during the physiological assays and to Roger Hutchings for the review of the English version of the manuscript.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.atech.2025.101042.

Fig. S1. Oxidative stress characterization in soybean plants submitted to stressful conditions during maturation. Hydrogen peroxide (H_2O_2) in non-stressed and stressed plants (A). Malondialdehyde acid (MDA) in non-stressed and stressed plants (B). MDA is a product of lipid peroxidation. Different letters indicate a significant difference ($P \leq 0.01$) by *t*-test. Error bars show the standard deviation from four samples.

Data availability

Data will be made available on request.

References

[1] D. Ajala-Luccas, J.P. Ribeiro-Oliveira, R.N. Teixeira, K.R. Ducatti, J.B. França-Neto, H.W.M. Hilhorst, E.A.A. da Silva, The seed–Seedling transition in commercial soybean cultivars with the presence of greenish seeds in the sample: a perspective from classical genetic parameters, Agronomy 13 (2023) 1966, https://doi.org/10.3390/AGRONOMY13081966, 2023, Vol. 13, Page 1966.

[2] V. Alexieva, I. Sergiev, S. Mapelli, E. Karanov, The effect of drought and ultraviolet radiation on growth and stress markers in pea and wheat, Plant. Cell. Environ. 24 (2001) 1337–1344, https://doi.org/10.1046/J.1365-3040.2001.00778.X.

[3] E.J. Ambrosano, R.T. Tanaka, H.A.A. Mascarenhas, B.van Raij, J.A. Quaggio, H. Cantarella, Leguminosas e oleaginosas, in: B.van. Raij, H. Cantarella, J. A. Quaggio, A.M.C. Furlani (Eds.), Recomendações De Adubação e Calagem Para o Estado de São Paulo, Instituto Agronômico de Campinas, Campinas, 1997, pp. 221–229 (IAC. Boletim Técnico, 100).

[4] P.A. Auler, G.M. Souza, M.R.G. da Silva Engela, M.N. do Amaral, T. Rossatto, M.G. Z. da Silva, C.M. Furlan, B. Maserti, E.J.B. Braga, Stress memory of physiological, biochemical and metabolomic responses in two different rice genotypes under drought stress: the scale matters, Plant Sci. 311 (2021) 110994, https://doi.org/10.1016/J.PLANTSCI.2021.110994.

[5] J.R. Bagateli, C.S. Dörr, L.O.B. Schuch, G.E. Meneghello, Productive performance of soybean plants originated from seed lots with increasing vigor levels, J. Seed Sci. 41 (2019) 151–159, https://doi.org/10.1590/2317-1545v41n2199320.

[6] C. Barboza da Silva, V.de J.M. Bianchini, A.D.de Medeiros, M.H.D.de Moraes, A. G. Marassi, A. Tannús, A novel approach for Jatropha curcas seed health analysis based on multispectral and resonance imaging techniques, Ind. Crops Prod. 161 (2021) 113186, https://doi.org/10.1016/J.INDCROP.2020.113186.

[7] C. Barboza da Silva, N.M. Oliveira, M.E.A. de Carvalho, A.D. de Medeiros, M. de Lima Nogueira, A.R. dos Reis, Autofluorescence-spectral imaging as an innovative method for rapid, non-destructive and reliable assessing of soybean seed quality, Sci. Rep. 11 (2021), https://doi.org/10.1038/s41598-021-97223-5.

[8] T.B. Batista, C.B. Mastrangelo, A.D. de Medeiros, A.C.P. Petronilio, G.R. Fonseca de Oliveira, I.L. dos Santos, C.A.C. Crusciol, E.A. Amaral da Silva, A reliable method to recognize soybean seed maturation stages based on autofluorescence-spectral imaging combined with machine learning algorithms, Front. Plant Sci. 13 (2022), https://doi.org/10.3389/fpls.2022.914287.

[9] V. Bianchini, J.M. de, G.M. Mascarin, L.C.A.S. Silva, V. Arthur, J.M. Carstensen, B. Boelt, C. Barboza da Silva, Multispectral and X-ray images for characterization of jatropha curcas L. seed quality, Plant Methods 17 (2021), https://doi.org/10.1186/s13007-021-00709-6.

[10] I. Cakmak, W.J.H. Cakmak, W.J. Horst, Effect of aluminium on lipid peroxidation, siiperoxide dismutase, catalase, and peroxidase activities in root tips of soybean (Glycine max), Physiol. Plant. 83 (1991) 463, https://doi.org/10.1111/j.1399-3054.1991.tb00121.x. –168.

[11] I. Cohen, S.I. Zandalinas, C. Huck, F.B. Fritschi, R. Mittler, Meta-analysis of drought and heat stress combination impact on crop yield and yield components, Physiol. Plant. 171 (2021) 66–76, https://doi.org/10.1111/PPL.13203.

[12] E. Contini, D.J.D. TalaminI, P.A. Vieira Junior, Cenário mundial de commodities: frango, soja e milho, Portal Embrapa (2013) [W W W Document]. URL, https://www.embrapa.br/busca-de-publicacoes/-/publicacao/968636/cenario-mundial-de-commodities-frango-soja-e-milho (accessed 11.21.23).
[13] L. Donaldson, Autofluorescence in plants, Molecules (2020), https://doi.org/10.3390/molecules25102393.
[14] L.A. Ebone, A. Caverzan, A. Tagliari, J.L.T. Chiomento, D.C. Silveira, G. Chavarria, Soybean seed vigor: uniformity and growth as key factors to improve yield, Agronomy 10 (2020) 545, https://doi.org/10.3390/AGRONOMY10040545, 2020, Vol. 10, Page 545.
[15] W.E. Finch-Savage, G.W. Bassel, Seed vigour and crop establishment: extending performance beyond adaptation, J. Exp. Bot. 67 (2016) 567–591, https://doi.org/10.1093/jxb/erv490.
[16] G.R. Fonseca de Oliveira, C.B. Mastrangelo, W.Y. Hirai, T.B. Batista, J.M. Sudki, A. C.P. Petronilio, C.A.C. Crusciol, E.A. Amaral da Silva, An approach using emerging optical technologies and artificial intelligence brings new markers to evaluate peanut seed quality, Front. Plant Sci. 13 (2022), https://doi.org/10.3389/fpls.2022.849986.
[17] França-Neto, J.B., Krzyzanowski, F.C., 2022. Metodologia Do Teste De Tetrazólio Em Sementes De Soja. Londrina.
[18] F. França-Silva, F.G. Gomes-Junior, C.H.Q. Rego, A.G. Marassi, A. Tannús, Advances in imaging technologies for soybean seed analysis, Journal of Seed Science 45 (2023) 1–16, https://doi.org/10.1590/2317-1545v45274098.
[19] P.A. Galletti, M.E.A. Carvalho, W.Y. Hirai, V.A. Brancaglioni, V. Arthur, C. Barboza da Silva, Integrating optical imaging tools for rapid and non-invasive characterization of seed quality: tomato (Solanum lycopersicum L.) and carrot (Daucus carota L.) as study cases, Front. Plant Sci. 11 (2020), https://doi.org/10.3389/fpls.2020.577851.
[20] J.I. García-Plazaola, B. Fernández-Marín, S.O. Duke, A. Hernández, F. López-Arbeloa, J.M. Becerril, Autofluorescence: biological functions and technical applications, Plant Sci. 236 (2015) 136–145, https://doi.org/10.1016/J.PLANTSCI.2015.03.010.
[21] A.A. Gitelson, Y. Gritz, M.N. Merzlyak, Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves, J. Plant Physiol. 160 (2003) 271–282, https://doi.org/10.1078/0176-1617-00887.
[22] W.P.C. Gomes, L. Gonçalves, C. Barboza da Silva, W.R. Melchert, Application of multispectral imaging combined with machine learning models to discriminate special and traditional green coffee, Comput. Electron. Agric. 198 (2022) 107097, https://doi.org/10.1016/J.COMPAG.2022.107097.
[23] International Seed Testing Association (ISTA), International Rules for Seed Testing, ISTA, Bassersdorf: Zürischstr, 2020. Edn.
[24] F.C. Krzyzanowski, F.B. França-Neto, A.A. Henning, A Alta Qualidade Da Semente De soja: Fator Importante Para a Produção Da Cultura CT136, EMBRAPA, 2018, pp. 1–24.
[25] H. Liu, A.J. Able, J.A. Able, Priming crops for the future: rewiring stress memory, Trends. Plant Sci. 27 (2022) 699–716, https://doi.org/10.1016/J.TPLANTS.2021.11.015.
[26] Lukić, N., Kukavica, B., Davidović-Plavšić, B., Hasanagić, D., Walter, J., 2020. Plant stress memory is linked to high levels of anti-oxidative enzymes over several weeks. https://doi.org/10.1016/j.envexpbot.2020.104166.
[27] S. Nakajima, H. Ito, R. Tanaka, A. Tanaka, Chlorophyll *b* reductase plays an essential role in maturation and storability of Arabidopsis seeds, Plant Physiol. 160 (2012) 261–273, https://doi.org/10.1104/PP.112.196881.
[28] N.M. Oliveira, A.D.de Medeiros, M. Nogueira, L. de, V. Arthur, T. Mastrangelo, A. de, C. Barboza da Silva, Hormetic effects of low-dose gamma rays in soybean seeds and seedlings: a detection technique using optical sensors, Comput. Electron. Agric. 187 (2021) 106251, https://doi.org/10.1016/J.COMPAG.2021.106251.
[29] G.P.de Pádua, J.de B. França-Neto, M.L.M. de Carvalho, O. Costa, F. C. Krzyzanowski, N.P. Da Costa, Tolerance level of green seed in soybean seed lots after storage, Rev. Bras. Sementes 29 (2007) 112–120, https://doi.org/10.1590/s0101-31222007000300015.
[30] G.P. Pádua, J. França-Neto, B. de, M.L.M. de Carvalho, F.C. Krzyzanowski, R. M. Guimarães, Incidence of green soybean seeds as a function of environmental stresses during seed maturation, Revista Brasileira de Sementes 31 (2009) 150–159, https://doi.org/10.1590/S0101-31222009000300017.
[31] S.W. Ritchie, J.J. Hanway, H.E. Thompson, How a soybean plant develops, Des Moines (1982).
[32] N. Sionit, P.J. Kramer, Effect of water stress during different stages of growth of soybean 1, Agron. J. 69 (1977) 274–278, https://doi.org/10.2134/agronj1977.00021962006900020018x.
[33] M. Staniak, E. Szpunar-Krok, A. Kocira, Responses of soybean to selected abiotic stresses—Photoperiod, temperature and water, Agriculture 13 (2023) 146, https://doi.org/10.3390/AGRICULTURE13010146.
[34] J.M. Sudki, G.R. Fonseca de Oliveira, A.D. de Medeiros, T. Mastrangelo, V. Arthur, E.A. Amaral da Silva, C.B. Mastrangelo, Fungal identification in peanuts seeds through multispectral images: technological advances to enhance sanitary quality, Front. Plant Sci. 14 (2023) 1112916, https://doi.org/10.3389/FPLS.2023.1112916/BIBTEX.
[35] R.N. Teixeira, W. Ligterink, J. França-Neto, B. de, H.W.M. Hilhorst, E.A.A. da Silva, Gene expression profiling of the green seed problem in Soybean, BMC Plant Biol. 16 (2016) 1–15, https://doi.org/10.1186/S12870-016-0729-0/FIGURES/10.
[36] USDA, Oilseeds: World Markets and Trade, United States Department of Agriculture, 2023 [W W W Document]URL, https://public.govdelivery.com/accounts/USDAFAS/subscriber/new.
[37] A.L. Villagómez-Aranda, ·.A.A. Feregrino-Pérez, ·.L.F. García-Ortega, ·.M. M. González-Chavira, ·.I. Torres-Pacheco, ·.R.G. Guevara-González, Activating stress memory: eustressors as potential tools for plant breeding, Plant Cell Rep. 41 (2022) 1481–1498, https://doi.org/10.1007/s00299-022-02858-x.
[38] X. Wang, H. Zhang, R. Song, M. Sun, P. Liu, P. Tian, P. Mao, S. Jia, Multiple omics datasets reveal significant physical and physiological dormancy in alfalfa hard seeds identified by multispectral imaging analysis, Crop. J. 11 (2023) 1458–1468, https://doi.org/10.1016/J.CJ.2023.03.003.
[39] S. Zhang, H. Zeng, W. Ji, K. Yi, S. Yang, P. Mao, Z. Wang, H. Yu, M. Li, Non-destructive testing of alfalfa seed vigor based on multispectral imaging technology, Sensors 22 (2022) 2760, https://doi.org/10.3390/S22072760.
[40] J. Zinsmeister, D. Lalanne, B. Ly Vu, B. Schoefs, J. Marchand, T.T. Dang, J. Buitink, O. Leprince, ABSCISIC ACID INSENSITIVE 4 coordinates eoplast formation to ensure acquisition of seed longevity during maturation in Medicago truncatula, Plant J. 113 (2023) 934–953, https://doi.org/10.1111/TPJ.16091.
[41] M. Zorato, F. de, S.T. Peske, C. Takeda, J.B. França Neto, Presença de sementes esverdeadas em soja e seus efeitos sobre seu potencial fisiológico, Revista Brasileira de Sementes 29 (2007) 11–19, https://doi.org/10.1590/S0101-31222007000100002.
[42] J.A. Davis. Elementary Survey Analysis, 1st ed., Prentice-Hall, Hoboken, NJ, USA, 1971.